Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1029/2018JD028657

Modeling of Nonlocal Thermodynamic Equilibrium Effects in the Classical and Principal Component-Based Version of the RTTOV Fast Radiative Transfer Model

Marco Matricardi[1], **Manuel López-Puertas**[2], **and Bernd Funke**[2]

[1]ECMWF, Reading, UK, [2]Instituto de Astrofisica de Andalucia, CSIC, Granada, Spain

Key Points:
- Development of a fast NLTE model for satellite nadir-sounding applications
- Extension of a principal component-based fast radiative transfer model to the simulation of NLTE effects
- The fast NLTE model can accurately simulate daytime shortwave satellite radiances but performs somewhat lesser at winter high latitudes

Correspondence to:
M. Matricardi,
marco.matricardi@ecmwf.int

Citation:
Matricardi, M., López-Puertas, M., & Funke, B. (2018). Modeling of nonlocal thermodynamic equilibrium effects in the classical and principal component-based version of the RTTOV fast radiative transfer model. *Journal of Geophysical Research: Atmospheres*, *123*, 5741–5761. https://doi.org/10.1029/2018JD028657

Received 12 MAR 2018
Accepted 11 APR 2018
Accepted article online 19 APR 2018
Published online 4 JUN 2018

Abstract The direct assimilation in 4D-Var of principal component (PC) scores derived from Infrared Atmospheric Sounding Interferometer (IASI) spectra has recently been demonstrated. To maximize the exploitation of the IASI instrument, a future step is to consider the extension of the PC approach to the extraction of information from the 4.3-μm CO_2-absorbing region. Shortwave IASI channels are currently underused compared to similar longwave channels because of day-night variations in data usability due to departures from local thermodynamic equilibrium (LTE). In this paper, we document the introduction of non-LTE (NLTE) effects in the PC-based version of the radiative transfer for TIROS operational vertical sounder (RTTOV) fast radiative transfer model (PC-RTTOV). The inclusion of NLTE effects in PC-RTTOV has required the development of a parameterized scheme that allows the fast computation of a NLTE correction to LTE radiances. The fast NLTE model is general enough to be applied to any sensor and can be utilized to add a fast and accurate NLTE correction to polychromatic LTE spectra computed by any general radiative transfer model, including RTTOV, which now incorporates the fast NLTE model developed in this study. The accuracy of the NLTE correction is such that daytime and nighttime radiances can be simulated to almost the same degree of accuracy. The comparison with IASI observations shows that the fast NLTE model presented here performs significantly better than the fast NLTE model incorporated in the previous version of RTTOV and also that improvements have to be made to the simulation of NLTE effects at winter high latitudes.

1. Introduction

The assimilation of high-spectral-resolution radiances measured by the Infrared Atmospheric Sounder Interferometer (IASI; Chalon et al., 2001) has produced a significant positive impact on forecast quality (Collard & McNally, 2009). The operational use of IASI radiances at numerical weather prediction centers is typically restricted to a selection of temperature sounding channels in the longwave region of the spectrum and to a limited number of ozone and humidity sounding channels. In principle, to exploit the full information content of IASI, the number of channels used in the assimilation could be increased to cover the full spectrum. Numerical weather prediction users are limited to assimilating less than the full IASI spectrum by the prohibitive computational cost, but it is also known that the independent information on the atmosphere contained in an IASI spectrum is significantly less than the total number of channels (Huang et al., 1992). There is thus a need to find a more efficient way of communicating the measured information to the assimilation system than simply increasing the number of channels.

Principal component analysis (PCA) is a classical statistical method for the efficient encapsulation of information from voluminous data (Joliffe, 2002). As such, it has been proposed as a solution to the above problem. In addition to reducing the dimension of the data, PCA can be applied to filter the random noise present in hyperspectral infrared observations (Antonelli et al., 2004). Principal component (PC)-based methodologies for the assimilation of high-resolution infrared sounder spectra include the use of radiances reconstructed from PCs (e.g., Collard et al., 2010) and the direct assimilation of PC amplitudes (also known as PC "scores") of radiances (e.g., Matricardi & McNally, 2013). A PC-based assimilation system requires a fast and accurate radiative transfer (RT) model that can calculate directly modeled equivalents of PC scores given first-guess fields of, for example, temperature, water vapor, ozone and surface parameters. Various fast RT models have been developed to do this (e.g., Havemann, 2006; Liu et al., 2006), including the PC-RTTOV (Matricardi, 2010)

fast RT model. PC-RTTOV—a PC-based version of the radiative transfer for TIROS operational vertical sounder (RTTOV) RT model (e.g., Matricardi et al., 2004)—has, for example, been used at the European Centre for Medium Range Weather Forecasts (ECMWF) to demonstrate the direct assimilation of IASI PC scores in a 4D-Var system that utilizes the full number of IASI channels in the spectral region between 645 and 2,000 cm^{-1} (i.e., the 5,421 channels comprised in IASI band 1 and band 2; Matricardi, 2016). To maximize the exploitation of the IASI instrument, a logical future step is to consider the extension of the PC approach to the extraction of information from IASI band 3 (i.e., the shortwave region between 2,000 and 2,760 cm^{-1}), which contains excellent temperature sounding channels.

Channels in IASI band 3 are currently underused compared to similar channels in the longwave region for a number of reasons, which include day-night variations in data usability due to departures from local thermodynamic equilibrium (LTE) and increased instrument noise (for details on the IASI instrument noise performance, see, e.g., Chalon et al., 2001). While IASI band 3 could benefit from PCA noise reduction, to avoid day/night sampling issues, the exploitation of shortwave PC data requires the introduction of non-LTE (NLTE) effects in the direct simulation of PC scores. NLTE occurs when the rate of collisional adjustment to the population of the energy states of a molecule is smaller than the rate of adjustment due to radiative processes (López-Puertas & Funke, 2015). For IASI-like satellite sounding applications, NLTE effects occur primarily during daytime above altitudes of ~40 km in the CO_2 spectral region at 4.3 μm where the main mechanism responsible for the population of the excited states is the absorption of the strong solar radiation field (e.g., López-Puertas & Taylor, 2001). If not accounted for, NLTE effects can introduce significant errors in the simulation of daytime top-of-the-atmosphere (TOA) IASI radiances. These errors can reach a maximum of ~10 K for the stratospheric sounding channels in the spectral region between 2,310 and 2,340 cm^{-1} and limit the use of the middle-to-upper tropospheric temperature sounding channels in the clean spectral region between 2,350 and 2,386 cm^{-1} where CO_2 represents the only absorbing specie. Note that for monochromatic radiances, unaccounted NLTE effects can result in errors that are significantly larger than 10 K.

In this paper, we document the extension of PC-RTTOV to the simulation of NLTE effects. PC-RTTOV exploits PCA to replace *n*-correlated variables with a smaller number of uncorrelated PCs. The PCs are calculated from a training set of *l* synthetic line-by-line (LBL) spectra of *n* radiances. The current version of PC-RTTOV (Matricardi, 2015) is based on 15,000 training spectra corresponding to a data set of diverse atmospheric situations over land and sea. The introduction of NLTE effects in the training LBL spectra (see Edwards et al., 1993, for details) would require the computation of an extremely large number of vibrational temperature profiles because we have to take into account a wide range of NLTE molecular states and we have to represent variations in the solar zenith angle. Methods for calculating accurate vibrational temperature profiles are complex and time consuming, and it would be materially impractical to use them to compute the sheer number of profiles needed for the training of PC-RTTOV. For this reason, following the approach used by DeSouza-Machado et al. (2007) and Chen et al. (2013), we have developed a computationally efficient algorithm that calculates the difference, ΔR^{NLTE}, between NLTE and LTE polychromatic LBL radiances. The fast NLTE algorithm is trained using a limited set of diverse vibrational temperature profiles and allows the rapid estimation of ΔR^{NLTE} for any atmospheric profile for any nadir viewing angle and solar zenith angle geometry. The NLTE algorithm described in this paper is general enough to be applied to any sensor and can be utilized to add a fast and accurate NLTE correction to the polychromatic LTE spectra computed by any general RT model, including RTTOV, which now incorporates the fast NLTE model developed in this study.

The outline of the paper is as follows. In section 2 we review the mechanism of NLTE emission, while in section 3 we describe the development of the fast NLTE algorithm for IASI. In section 4 we discuss the accuracy of the fast NLTE algorithm using reference LBL spectra and IASI observations. In the same section, we compare the performance of our fast NLTE algorithm to the performance of the fast NLTE algorithm incorporated in the previous version of the RTTOV model (i.e., RTTOV-11, see Saunders et al., 2013, for details) and discuss the possible origin of larger biases in the NLTE simulations at winter high latitudes. In section 5 we describe the training of PC-RTTOV in presence of NLTE effects and give an assessment of the accuracy of the PC-RTTOV NLTE calculations using reference LBL spectra and IASI observations. Finally, in section 6 we present a summary of the results and discuss possible future directions.

2. NLTE Emission

The LBL radiances used to train the PC-RTTOV fast RT model are currently computed under the assumption that all energy states of the molecules included in the LBL computations are in LTE. In a regime of LTE, the distribution of molecules between two states (typically the ground state and an excited state) is given by Boltzmann's law:

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-\frac{\Delta E}{kT_k}} \tag{1}$$

where T_k is the local kinetic temperature, k is the Boltzmann constant, $\Delta E = E_1 - E_0$ is the energy difference between the excited and the ground state, and g_i is the statistical weight of level i. The concept of LTE applies to all energy modes (e.g., translational, rotational, and vibrational modes). For a gas medium where radiation is neglected, collisions between gas molecules are sufficient to maintain a Boltzmann distribution of population numbers. Conversely, in a medium with no molecular collisions where only radiative processes take place, the population numbers can differ significantly from the values given by equation (1). In some situations, nonthermal processes (e.g., chemical reactions) can dominate over collisional and radiative processes and one can say that in general, the population of the energy states is determined by the competing effects of collisional, radiative and nonthermal processes (see López-Puertas & Taylor, 2001, for more details).

In a classical situation (i.e., not involving nonthermal processes), LTE condition is established when the rate of collisional adjustment to the population of the energy states is much larger than the rate of adjustment due to radiative processes. Although collisional processes occur much more frequently than radiative processes, the frequency of collisions varies inversely with air density and we therefore expect that LTE breaks down in the upper atmosphere. This is not the case for the translational modes, that is, those associated with molecular kinetic energy. For these modes, only a few molecular collisions are needed to bring the populations of the energy states into equilibrium. Translational energy states are in LTE at all altitudes below the base of the exosphere (i.e., the upper limit of the Earth's fluid atmosphere), and consequently, the concept of kinetic temperature (and Maxwell's distribution of molecular velocities) is valid at all atmospheric levels of practical interest for terrestrial remote sensing applications. Similar to translational energy states, rotational energy states require a limited number of collisions to be brought into equilibrium and, unless the atmosphere is too rarefied, are normally considered to be in LTE. Conversely, vibrational energy states may require a large number of collisions to be brought in equilibrium. If local thermal equilibrium cannot be maintained through collision in the gas, the vibrational populations cannot be described by a Boltzmann distribution and the corresponding energy states are said to be in NLTE. However, equation (1) can still give the correct population numbers if the local kinetic temperature is replaced by a local excitation (or vibrational) temperature, which may be different for different energy levels. This is also true for nonclassical situations involving nonthermal processes (López-Puertas & Taylor, 2001).

For satellite nadir-sounding applications, we are primarily concerned with NLTE effects above altitudes of ~40 km in the CO_2 spectral region at 4.3 μm. Here LTE breaks down because the main mechanism responsible for the population of the excited states is the absorption of the strong solar radiation field. If not accounted for, NLTE effects at 4.3 μm can introduce large errors in the simulation of daytime TOA radiances. NLTE also affects the CO_2 spectral region at 15 μm because above altitudes of ~ 70 km the frequency of the molecular collisions is not high enough to supply the energy lost by radiative processes (i.e., the spontaneous emission of photons). However, NLTE effects in this region have a negligible impact on TOA radiances.

As shown by Edwards et al. (1993), the introduction of NLTE effects in LBL models requires modifications to the molecular absorption coefficient, the source function, and the vibrational partition function. These modifications involve the specification of the ratio of populations between the NLTE and LTE cases for a given state. Because the ratio of populations can be expressed in terms of vibrational and kinetic temperature, the LBL calculation of NLTE spectra requires the knowledge of the altitude dependent vibrational temperature of the states that are not in equilibrium.

3. The Fast NLTE Algorithm for IASI

As discussed in the introduction, the extension of PC-RTTOV to the simulation of NLTE effects in the CO_2 4.3-μm spectral region requires the computation of supplemental LBL training spectra incorporating the effect of

Table 1
The Predictors Used in the Fast NLTE Algorithm

Predictor number	Predictors
1	constant
2	$\cos(\theta_\odot)$
3	$(\cos(\theta_\odot))^{0.5}$
4	$\cos(\theta_\odot)\ \sec(\theta)$
5	$(\cos(\theta_\odot)\ \sec(\theta))^2$
6	$\cos(\theta_\odot)\ T_1^{av}$
7	$\cos(\theta_\odot)\ T_2^{av}$
8	$\sec(\theta)\ T_1^{av}$
9	$\sec(\theta)\ T_2^{av}$

NLTE emission. To represent the dependence of NLTE effects on daily variations of solar radiation, NLTE spectra have to be computed for N different values of the solar zenith angle. If we tentatively assume a value of 10 for N, given the current number of 15,000 LTE training spectra, the NLTE training of PC-RTTOV would require the computation of ~150,000 vibrational temperature profiles for each of ~50 different CO_2 energy levels. The task of computing this number of accurate vibrational temperature profiles is computationally prohibitive. Consequently, following DeSouza-Machado et al. (2007) and Chen et al. (2013), we have developed a computationally efficient algorithm that can be utilized to apply a fast NLTE correction, ΔR_i^{NLTE}, to LTE LBL polychromatic radiances. A key feature of the fast NLTE algorithm is that it does not require explicit knowledge of the vibrational temperatures; these are only used for the training of the algorithm.

The fast NLTE algorithm developed in this paper calculates the difference $\Delta R_i^{NLTE} = R_i^{NLTE} - R_i^{LTE}$ between NLTE and LTE LBL radiances in IASI channel i through a regression relation in which a set of profile-dependent predictors is governed by a set of channel-dependent coefficients:

$$\Delta R_i^{NLTE} = \sum_{j=1}^{n} X_{i,j}\, b_j \tag{2}$$

In equation (2), b_j are the predictors and $X_{i,j}$ are coefficients that have been estimated by fitting equation (2) to a data set of ΔR_i^{NLTE} LBL radiances between 2,200 cm^{-1} (i = 6,221) and 2,400 cm^{-1} (i = 7021). The predictors used in our regression model are tabulated in Table 1. It can be seen that the predictors consist of various combinations of the solar zenith angle ($\theta_\odot$), the sensor zenith angle (θ), and the average kinetic temperatures in two broad atmospheric layers above 50 hPa (T_1^{av} is the average temperature between 0.005 and 0.2 hPa, and T_2^{av} is the average temperature between 0.2 and 50 hPa). The predictors reflect the strong dependence of NLTE effects on solar zenith angle and the fact that deviations of vibrational temperatures from kinetic temperatures start becoming significant in the stratospheric layers above ~50 hPa. The secant of the sensor zenith angle describes the off-nadir angular dependence of NLTE TOA radiance observations. The predictors have been determined using a forward stepwise selection; that is, starting from a model with no predictors (only the intercept), we have chosen the single variable model that resulted in the lowest residual sum of squares. We have successively added more variables to the model based on the improvement of the statistics of the residual sum of squares. The process has been repeated until the statistics of the residual sum of squares stopped improving.

It should be noted that equation (2) is similar to the one utilized in the fast NLTE model developed by DeSouza-Machado et al. (2007). However, although DeSouza-Machado et al. (2007) also use predictors based on combinations of the solar zenith angle, sensor zenith angle, and average temperature, the functional form of our regression model is substantially different. In addition, we use a larger number of predictors (i.e., 9 versus 6) and have included the T_1^{av} term, which is absent in the model by DeSouza-Machado et al. (2007).

Regarding the predictors used in the fast NLTE algorithm, we should note that vibrational temperatures depend on CO_2 concentrations. The CO_2 profiles utilized for the computation of the training vibrational temperatures cover the range of atmospheric conditions represented by the six equinox (April) cases listed in Table 2 by Funke et al. (2012). Vibrational temperatures can be significantly affected by CO_2 variations in the upper atmosphere. Crucially, however, vibrational temperatures are little affected by CO_2 variations in the atmospheric regions that give the bulk contribution to the outgoing radiances measured by IASI. To this end, we have carried tests that show that the sensitivity of the NLTE correction to a 10% perturbation of the CO_2 concentrations is typically less than 0.1 K. For this reason, our NLTE regression model does not include CO_2-dependent predictors. Although vibrational temperatures do not depend directly on the abundances of gas species (CO_2 is an exception), it should be noted that the photodissociation of O_3 produces the excited oxygen atom $O(^1D)$ which is an important NLTE excitation source of the CO_2 ($\nu 3 = 1$) level (i.e., the fundamental band emission near 4.3 μm) at altitudes of 50–70 km. Because at these altitudes O_3 constrains $O(^1D)$ concentrations, O_3 is an important parameter for the calculation of NLTE populations. Although this suggests that

the choice of predictor variables employed by the fast NLTE algorithm should include O_3 terms, these are actually absent in the regression model because they explain very little of the variance in NLTE TOA radiances. We can explain this result by noting that the ozone profiles utilized for the computation of the training vibrational temperatures cover a very narrow range of atmospheric conditions, that is, those represented by the six equinox (April) cases listed in Table 2 of Funke et al. (2012). The lack of an adequate representation of effects due to enhanced O_3 variability is an obvious limitation of the fast NLTE model because it can potentially affect the accuracy of the simulated NLTE radiances. We shall return to this aspect in section 4.2.

To train the fast NLTE algorithm, we have computed a database of high-resolution monochromatic NLTE LBL radiances, R^{NLTE}, for a data set of 48 diverse atmospheric profiles extending from the surface to an altitude of 120 km. For each profile we have computed NLTE radiances for five sensor viewing angles (i.e., the angles for which the secant has equally spaced values from 1 to 2) and 12 solar zenith angles (i.e., 0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, 85°, 87°, and 90°) utilizing vibrational temperature profiles for the NLTE states of CO_2. This corresponds to a total number of 2,880 NLTE LBL training spectra to which we have added a further 240 LTE LBL spectra, R^{LTE}, to form the $\Delta R_i^{non-LTE}$ differences used in the regression (we only require 240 LTE LBL spectra because these spectra do not depend on the solar zenith angle). Each monochromatic LBL spectrum has been convolved with the appropriate instrument spectral response function to obtain the IASI polychromatic radiances in each channel *i*. Because the fast NLTE model presented in the paper is intended for clear-sky applications, the training LBL spectra do not include the contribution of scattering of solar radiation by aerosols and clouds. It should also be mentioned that since NLTE channels are sensitive to emission in the upper atmosphere, the effects of clouds can be neglected. To calculate the vibrational temperature profiles, we have used the latest version of the GRANADA NLTE population algorithm (Funke et al., 2012). Note that compared to the published version of the algorithm, we have applied modifications that include the use of collisional rates recently derived from Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) spectra (Jurado-Navarro et al., 2015) and the update of two of the above collisional rates (see pp. 6084–6085 of Jurado-Navarro et al., 2016, for details).

To broaden the scope of the study, the NLTE regression algorithm described in this paper has been used to produce a supplementary set of regression coefficients based on a preexisting database of vibrational temperatures (the same vibrational temperatures utilized by DeSouza-Machado et al., 2007) calculated with an older version of the GRANADA NLTE population algorithm (Funke et al., 2002). This will allow us to assess the relative merits of the old and new version of the GRANADA algorithm. It should be noted that although the derivation of the supplementary set of regression coefficients has involved the use of the same 48 atmospheres utilized to derive the primary set of regression coefficients, vibrational temperatures were only available for seven values of the solar zenith angle (i.e., 0°, 40°, 60°, 80°, 85°, 90°, and 120°). This means that the training has been carried out using 1,440 NLTE spectra. With regard to the supplementary set of regression coefficients, we should note that the previous version of the RTTOV model (i.e., RTTOV-11, Saunders et al., 2013) incorporated a fast NLTE model (Chen et al., 2013) based on exactly the same underlying database of vibrational temperatures and training atmospheres. It is instructive to note that the regression scheme employed by the RTTOV-11 fast NLTE model (hereafter referred to as the old fast NLTE model) is significantly different from the regression scheme employed by the fast NLTE model (hereafter referred to as the new fast NLTE model) presented in this paper. More specifically, while the predictors employed by the new fast NLTE model allow the computation of global regression coefficients, those employed by the old fast NLTE model only allow the computation of coefficients at discrete points, that is, for a couple of sensor and solar zenith angles. In other words, in the old fast NLTE model sensor and solar zenith angle are not independent variables in the regression. To compute the NLTE correction for a general value of the sensor and solar zenith angle, the old fast NLTE model uses a bilinear interpolation operator.

The training data set of NLTE and LTE LBL spectra has been calculated using the Line-By-Line Radiative Transfer Model (LBLRTM; Clough et al., 2005) developed by Atmospheric and Environmental Research (AER). More specifically, we have used LBLRTM version 12.2 and the associated mandatory list of molecular parameters *aer_v_3.2*. The *aer_v_3.2* molecular database is largely drawn from HITRAN_2008 (Rothman et al., 2009) with AER-customized updates. In our LBL calculations we have included 17 atmospheric constituents, that is, H_2O, O_3, CO_2, CH_4, N_2O, CO, NO_2, SO_2, NO, N_2, O_2, HNO_3, OCS, CCl_4, CF_4, CCl_3F (CFC-11), and CCl_2F_2 (CFC-12), and have used the same vertical profiles of pressure, temperature, and gas abundances

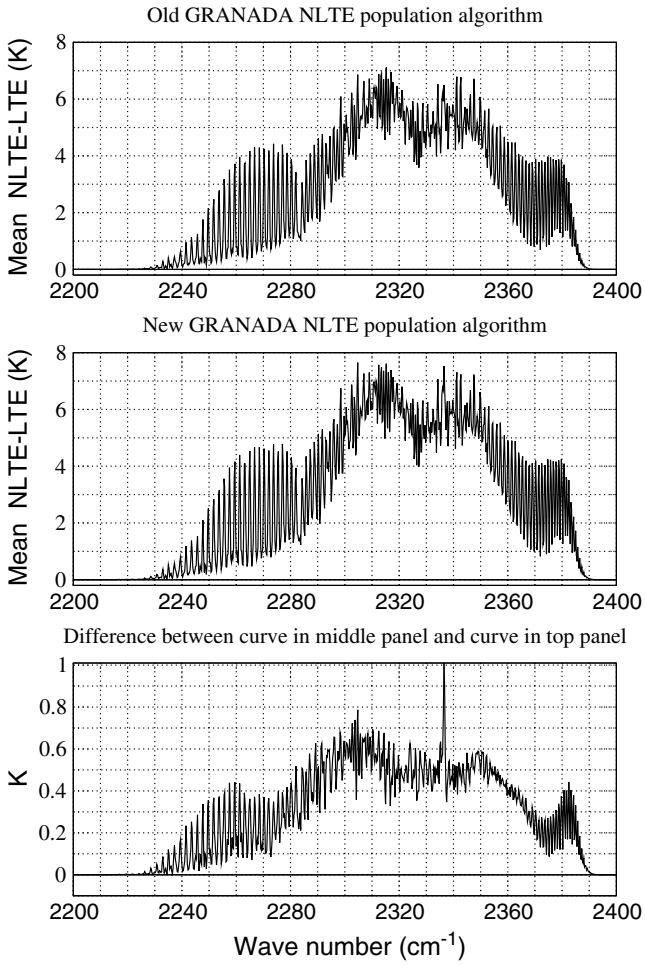

Figure 1. The statistics of the difference between NLTE line-by-line and LTE line-by-line computations. The top panel shows results for Infrared Atmospheric Sounding Interferometer obtained using vibrational temperatures computed with the old GRANADA population algorithm (Funke et al., 2002), whereas the middle panel shows results for Infrared Atmospheric Sounding Interferometer obtained using vibrational temperatures computed with the new GRANADA population algorithm (this work). The lower panel shows the difference between the curve plotted in the middle panel and the curve plotted in the top panel. LTE = local thermodynamic equilibrium; NLTE = non-LTE.

utilized in the computation of the vibrational temperatures. Regarding the vibrational temperatures used in the NLTE LBL computations, the GRANADA data sets comprises vibrational temperatures corresponding to vibrational transitions for 57 CO_2 states (27 for the main $^{12}C^{16}O_2$ isotopologue, 9 for the $^{13}CO_2$ isotopologue, 9 for the $^{16}O^{12}C^{18}O$ isotopologue, and 12 for the $^{16}O^{12}C^{17}O$ isotopologue), whereas LBLRTM allows the use of vibrational temperatures only for 55 CO_2 states. However, the two states (12211 and 12212, main isotope) not featured in the LBLRTM NLTE computations correspond to high energy levels belonging to weak transitions and hence of negligible importance.

To document the expected magnitude of NLTE effects in the CO_2 spectral region at 4.3 μm, in Figure 1 we have plotted the mean value of the difference between NLTE LBL and LTE LBL IASI brightness temperature spectra for the atmospheric situations used for the training of the fast NLTE model. More specifically, in the upper panel we have plotted results obtained using vibrational temperatures computed with the old GRANADA NLTE population algorithm, whereas in the middle panel we have plotted results obtained using vibrational temperatures computed with the new GRANADA NLTE population algorithm. As discussed previously, the old and the new GRANADA data sets comprise vibrational temperature profiles for 6 and 12 values of the daytime solar zenith angle, respectively. Consequently, for consistency reasons, the statistics have been computed using the 1,440 spectra that correspond to common values of the solar zenith angle (i.e., the six daytime values used in the old GRANADA data set). Figure 1 shows that NLTE effects interest the spectral region between 2,230 and 2,387 cm^{-1}. This spectral region comprises IASI channels with weighting functions peaking above ~400 hPa. The maximum deviation from LTE occurs at wave numbers between 2,310 and 2320 cm^{-1} (i.e., channels peaking between 9 and 50 hPa). Although not shown in Figure 1, the magnitude of NLTE effects increases with decreasing solar zenith angle. In the lower panel of Figure 1 we have plotted the difference between the curve shown in the middle panel and the curve shown in the top panel. The curve plotted in the lower panel of Figure 1 shows that the magnitude of the NLTE effects computed using the new GRANADA population algorithm is consistently larger than the magnitude of the NLTE effects computed using the old GRANADA population algorithm. This result can be explained by noting that vibrational temperatures computed by the new GRANADA algorithm are generally larger than vibrational temperatures computed by the old GRANADA algorithm. Differences between vibrational temperatures are illustrated in Figure 2 where we have plotted the mean value of the difference between old and new vibrational temperatures for a sample that comprises 288 cases, that is, 48 training profiles and six solar zenith angles (0°, 40°, 60°, 80°, 85°, and 90°). Results are shown for CO_2 level 00111 originating in the 4.3-μm fundamental band, CO_2 level 01111 originating in the 4.3-μm first hot band, and CO_2 levels 10012, 02211, and 10011 originating in the 4.3-μm second hot bands. The curves in Figure 2 are labeled using the HITRAN (e.g., Rothman et al., 2009) two-digit notation; that is, the first and second digits indicate the molecule and the isotopologue number, respectively (CO_2 is molecule number 2 in HITRAN).

4. Performance of the Fast NLTE Algorithm for Simulation of IASI NLTE Radiances

4.1. Validation Against Reference LBL Spectra

The accuracy of the regression algorithm developed to predict NLTE effects in the IASI shortwave channels can be assessed by comparing the radiances predicted by the fast NLTE model with LBL model

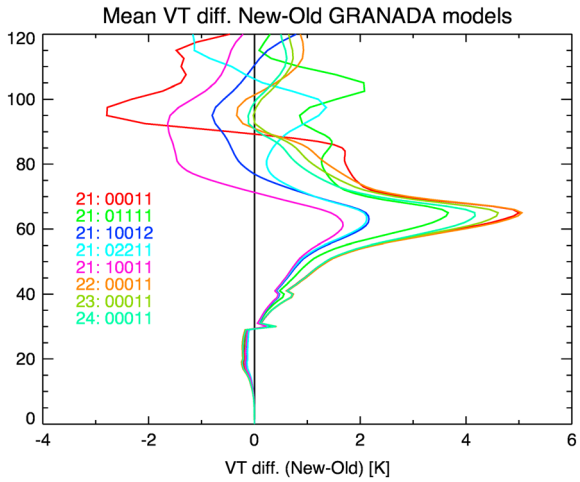

Figure 2. The mean value of the difference between vibrational temperatures computed using the old and new versions of the GRANADA nonlocal thermodynamic equilibrium population algorithm. The curves show results for the CO_2 (00011), (01111), (10012), (02211), and (10011) levels of the main isotopologue and the CO_2 (00011) level of the second, third, and fourth isotopologue, respectively.

equivalents. An initial test has been carried out using the 48 training profiles. Although these profiles form a dependent set, they are diverse enough to provide a realistic estimate of the error. To this end, the fast NLTE model has been used to calculate NLTE radiances, $R_i^{\mathrm{NLTEFAST}} = R_i^{\mathrm{LBL}} + \Delta R_i^{\mathrm{NLTEFAST}}$, for the 48 atmospheres, 5 viewing angles, and 12 solar zenith angles providing a total of 2,880 cases. The results have been then compared with the original LBLRTM NLTE radiances, R_i^{NLTELBL}. The means and the standard deviations of the difference statistic, $R_i^{\mathrm{NLTEFAST}} - R_i^{\mathrm{NLTELBL}}$, are shown in the upper left and lower left panels of Figure 3, respectively. It can be seen that the means of the difference statistics $R_i^{\mathrm{NLTEFAST}} - R_i^{\mathrm{NLTELBL}}$ are a negligible fraction of the standard deviations. The standard deviations tend to increase with the altitude of the peak of the channel weighting function and typically do not exceed 0.15 K.

Following Rayer (1995), we have carried out a cross-validation test. This involves the calculation of new regression coefficients with one of the training profiles omitted. The new regression coefficients can then be used to compute the fast NLTE radiances for the 5 viewing angles and the 12 solar zenith angles of the omitted profile, and a comparison can be made with the LBL NLTE radiances as before. This procedure can be applied to all 48 profiles by dropping each in turn. By doing so, each profile is independent of the regression coefficients. This is quite a stringent test because the training data set contains extreme profiles, and dropping one of them is therefore likely to make the task of predicting that particular profile much more difficult than would be the case for one chosen at random from the population at large. Results for the cross-validation test have been plotted in the upper right and lower right panel of Figure 3. They show that fast NLTE model errors are generally robust. The passage from the dependent set to the cross-validation test case results in a modest increase of the standard deviations, typically less than 30%. The means are virtually unchanged.

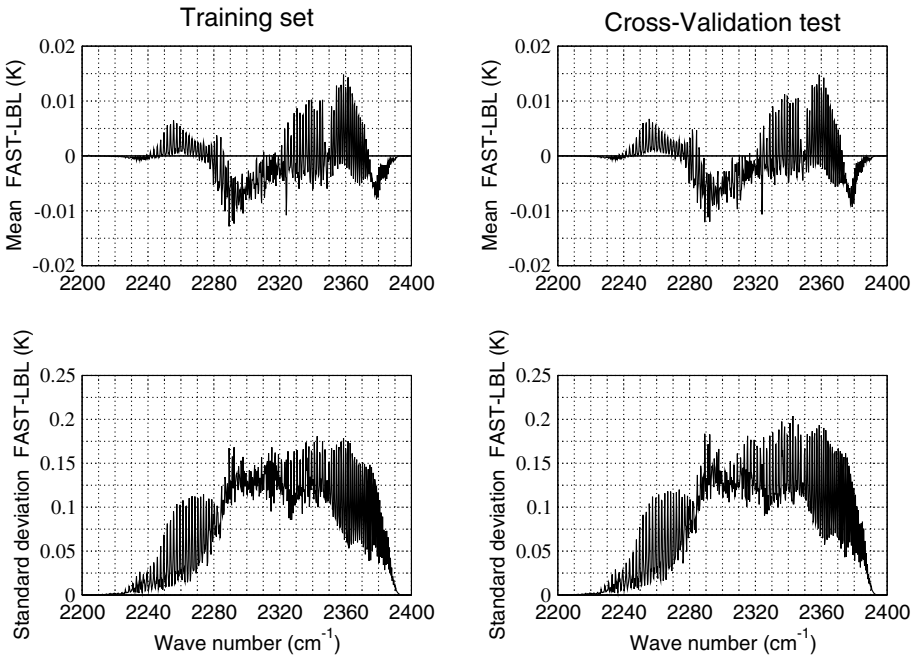

Figure 3. The mean value (upper panels) and the standard deviation (lower panels) of the difference between fast model and line-by-line (LBL) model nonlocal thermodynamic equilibrium radiances for the training set (left panels) and the cross validation test (right panels).

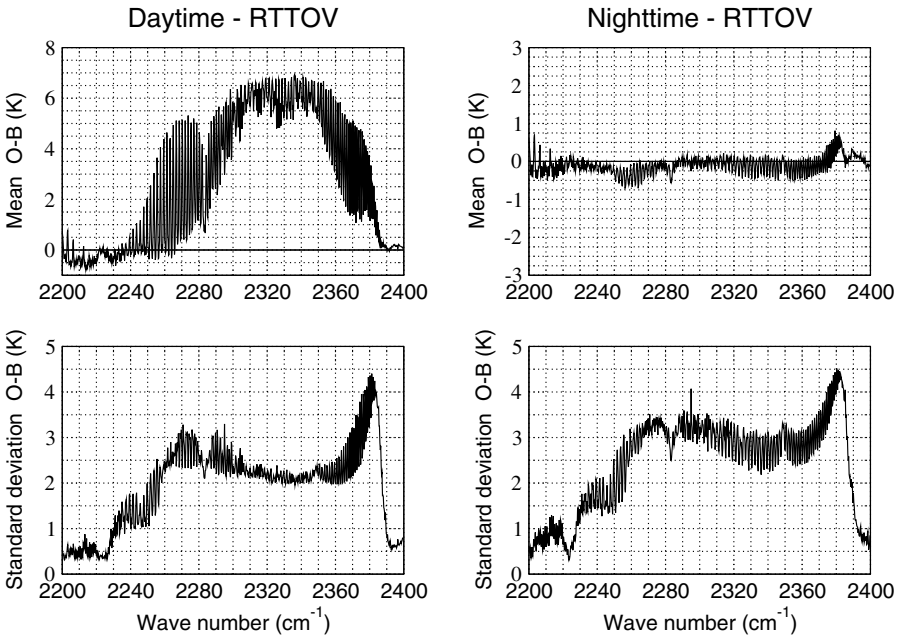

Figure 4. The mean value (bias; upper panels) and the standard deviation (lower panels) of the difference between observed radiances and radiances computed using RTTOV. The left panels show results obtained during daytime, whereas the right panels show results obtained during nighttime. RTTOV simulations do not include nonlocal thermodynamic equilibrium effects. RTTOV = radiative transfer for TIROS operational vertical sounder.

4.2. Validation Against Observations

Results discussed in section 4.1 represent the ability of the fast NLTE model to reproduce accurate LBL simulations of NLTE effects. In this section, we present results obtained by comparing IASI spectra measured over sea with simulations carried out using the classical RTTOV model. More precisely, we have merged the new fast NLTE model into RTTOV and have compared IASI clear brightness temperature observations (O) between 2,200 and 2,400 to RTTOV simulations (B) carried out using the new fast NLTE model and the old fast NLTE incorporated in RTTOV-11 (see section 3). To identity IASI clear channels, we have used the ECMWF cloud detection algorithm (McNally & Watts, 2003), which identifies clear channels at a particular location rather than clear locations. This means that the size of the data sample varies with the sensitivity of the channel to clouds. The sample amounts to tens of thousands of observations for channels less sensitive to clouds (i.e., channels peaking at middle and high altitudes, which include NLTE channels) to a few thousand of observations for channels more sensitive to clouds (i.e., channels peaking at low altitudes or near the surface).

As input to the RTTOV simulations, we have used global background (i.e., short-range forecast) fields of temperature, water vapor, and ozone from the ECMWF model during the period 10–20 November 2014. The background fields used in this study have been generated using version 41r2 of the ECMWF model (also known as cycle 41r2 of the ECMWF Integrated Forecast System). The horizontal resolution of the forecasts is ~9 km on 137 vertical levels. The ECMWF model uses a hybrid vertical coordinate system with pressure coordinates that follow the orography of the terrain in the lower and middle troposphere and fixed pressure coordinates in the stratosphere above about 60 hPa. Pressure at the model top is 0.01 hPa (~85 km).

RTTOV simulations including LTE only effects have been used to produce the results shown in Figure 4 where we have plotted the means (upper panels) and the standard deviations (lower panels) of the difference statistics, $O - B$, during daytime (left panels) and nighttime (right panels). The upper left panel of Figure 4 clearly shows that during daytime there is a large positive bias that is absent during nighttime. The daytime bias curve in Figure 4 is similar in shape and magnitude to the simulated curve plotted in, for example, the upper panel of Figure 1 indicating that the bulk of the bias originates from the absence of NLTE processes in the radiance simulations. The reader will note that standard deviations during

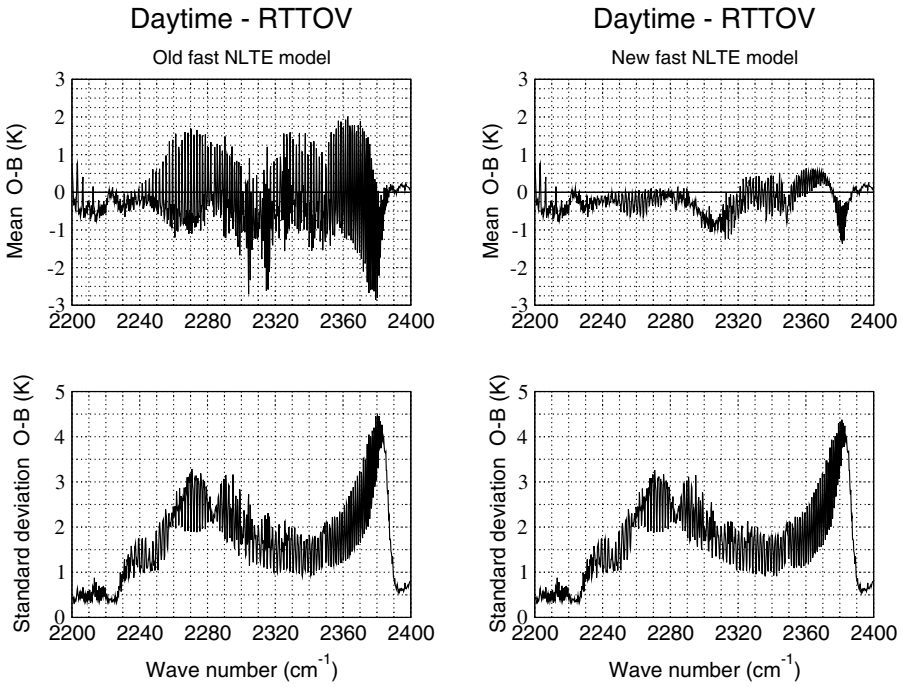

Figure 5. The mean value (bias; upper panels) and the standard deviation (lower panels) of the difference between observed daytime radiances and radiances computed using RTTOV. RTTOV simulations include NLTE effects computed using the old fast NLTE model (left panels) and the new fast NLTE model (right panels). RTTOV = radiative transfer for TIROS operational vertical sounder; NLTE = nonlocal thermodynamic equilibrium.

daytime are lower than standard deviations during nighttime. This can be explained by noting that in the spectral region plotted in Figure 4, standard deviations are dominated by the IASI instrument noise, which, when expressed in brightness temperature units, decreases with the increased scene temperatures observed during daytime.

4.2.1. Results for the Old and New Fast NLTE Models

Daytime results obtained introducing NLTE effects in the RTTOV simulations are shown in the left and right panels of Figure 5 for the old fast NLTE model and the new fast NLTE model, respectively. It is evident that the introduction of NLTE effects in the RTTOV simulations greatly reduces the daytime biases and, to an extent, standard deviations. Both fast NLTE models produce daytime results that are much closer to results obtained during nighttime. However, mean *O-B* departures computed using the old fast NLTE model have substantially larger biases. Here we should note that the statistics of the *O-B* departures has been computed using an identical data sample and that simulations have been carried out using the same RTTOV fast transmittance model. This means that the larger biases observed for the old fast NLTE model can only be explained by larger biases in the simulation of the fast NLTE correction term, ΔR_i^{NLTE}.

In section 3 we have already noted that the old and the new fast NLTE models are based on different vibrational temperatures (i.e., they use vibrational temperatures calculated using the old and new version of the GRANADA population algorithm, respectively) and that they employ different algorithms to carry out the regression-based prediction of the NLTE radiances. An in-depth study of the performance of the old fast NLTE model is outside the scope of this paper. However, because the new fast NLTE model can optionally use regression coefficients based on the old GRANADA vibrational temperatures (see section 3), we can at least use this capability to try to understand whether the larger biases for the old fast NLTE model are due to the use of the old GRANADA vibrational temperatures.

Results obtained using the new fast NLTE model based on the old GRANADA vibrational temperatures are indicated by the red curves in Figure 6. These results strongly suggest that the use of the old GRANADA vibrational temperatures cannot explain the behavior exhibited by the old fast NLTE model. If this were the case, we would expect that the shape and magnitude of the red curve in the upper panel of Figure 6 would match very closely the shape and magnitude of the curve shown in the upper left panel of Figure 5. Thus, unless

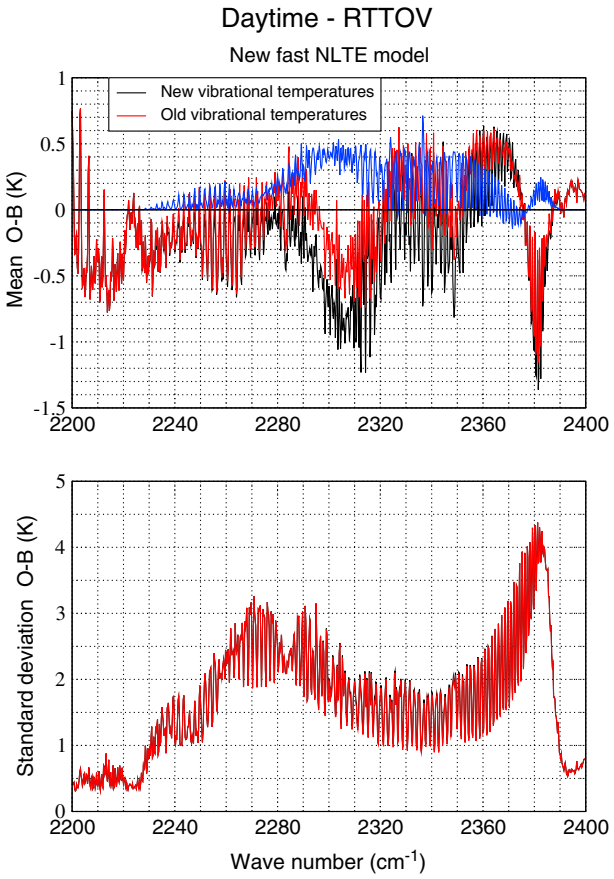

Figure 6. The mean value (bias; upper panel) and the standard deviation (lower panel) of the difference between observed daytime radiances and radiances computed using RTTOV. The black curve shows results obtained using the new fast NLTE model based on the new GRANADA vibrational temperatures, whereas the red curve shows results obtained using the new fast NLTE model based on the old GRANADA vibrational temperatures. The blue curve in the top panel is the difference between the red curve and the black curve in that same panel. RTTOV = radiative transfer for TIROS operational vertical sounder; NLTE = nonlocal thermodynamic equilibrium.

simulations are affected by errors in the computation of the training NLTE LBL radiances, we can conclude that the bulk of the larger biases for the old fast NLTE model is very likely caused by the regression scheme utilized to compute the NLTE correction to the radiances.

The availability of optional regression coefficients offers us the possibility of using the new fast NLTE model to evaluate the relative merits of the old and new GRANADA population algorithms. To this end, the red curves shown on the graphs in Figure 6 have been superimposed with black curves representing results for the new fast NLTE model based on the new GRANADA vibrational temperatures. Note that the black curves in Figure 6 are identical to those plotted in the right panels of Figure 5.

In section 3, we have shown that LBL simulations carried out using the new GRANADA vibrational temperatures produce NLTE TOA radiances that are warmer than those produced by simulations carried out using the old GRANADA vibrational temperatures. It is gratifying to see that biases plotted in the upper panel of Figure 6 exhibit a similar behavior (i.e., the black curve tends to lie consistently below the red curve). This attests to the robustness of the simulations carried out by the new fast NLTE model. This point is further illustrated by the close resemblance between the curve plotted in the lower panel of Figure 1 and the curve representing the difference between the red and the black curve in the upper panel of Figure 6 (i.e., the blue curve in Figure 6). In many spectral regions, the new fast NLTE model based on the old GRANADA vibrational temperatures has a lower bias. It is difficult, however, to conclude that the old vibrational temperatures are more accurate than the new ones. Because biases are generally small, there is no obvious way to distinguish between systematic errors from the atmospheric state, the RTTOV forward model (including spectroscopic errors) and the fast NLTE model. Thus, the new vibrational temperatures might still be more accurate and yet result in larger residuals due to the competing effect of biases from other sources. This said, although not immediately apparent from Figure 6, standard deviations for the new fast NLTE model based on the new vibrational temperatures are marginally higher than standard deviations for the model based on the old vibrational temperatures. These findings warrant further investigation in future studies.

4.2.2. Considerations on Fast NLTE Model Errors at Winter High Latitudes

We can gain more insight into the nature of the residuals of the new fast NLTE model with respect to IASI observations by plotting them in terms of binned departures. More precisely, for each channel in the region between 2240 and 2385 cm^{-1}, *O-B* departures have been averaged over solar zenith angles intervals of 2.5°. We have then partitioned the absolute values of the biases in bins of equal width (i.e., 0.5 K) and have plotted them in the stacked bar chart shown in Figure 7. In this figure, each shaded region corresponds to a different solar zenith angle interval. Note that results have been obtained using the "standard" regression coefficients, that is, those based on the new GRANADA vibrational temperatures. Figure 7 shows that, while smaller biases tend to be equally distributed among all solar zenith angles, larger biases tend to be concentrated in an increasingly narrow range of larger solar zenith angles. For instance, 95% of the biases between 3.5 and 4 K are found for solar zenith angles between 82.5° and 87.5°. Regarding the results shown in Figure 7, we note that IASI daytime observations corresponding to large values of the solar zenith angle occur during the transit at higher latitudes. For the boreal winter conditions considered so far, solar zenith angles greater than, for example, 70°, correspond to latitudes higher than 45°N and 70°S. Note that the solar zenith angles for which we have the larger biases correspond almost exclusively to transits in the Northern Hemisphere.

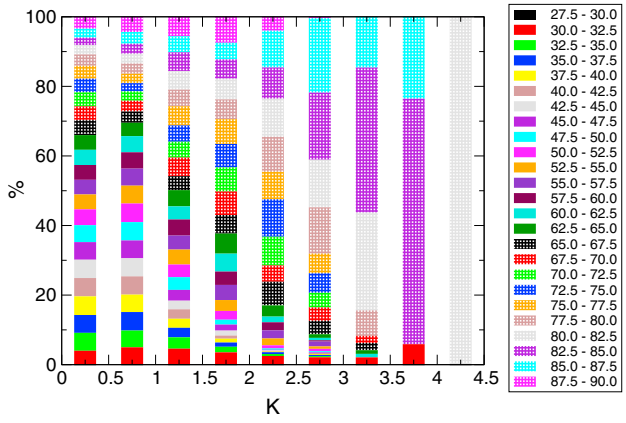

Figure 7. The stacked bar chart of binned means of differences between observed shortwave Infrared Atmospheric Sounding Interferometer radiances and Infrared Atmospheric Sounding Interferometer radiances computed using the old fast nonlocal thermodynamic equilibrium model expressed as a function of binned solar zenith angles.

Although results shown in Figure 7 provide a useful insight into the distribution of the biases, they do not provide any indication of the physical processes whereby biases increase at high latitudes. A candidate for such mechanism was proposed by DeSouza-Machado et al. (2007) who applied their fast NLTE model to Atmospheric Infrared Sounder (AIRS) spectra. They suggested that the enhanced radiance biases are caused by enhanced biases in the ECMWF stratospheric/lower-mesospheric temperature fields at high latitudes during winter. For this hypothesis to be corroborated there must be a consistency between results obtained in the 4.3-μm (shortwave) and 15-μm (longwave) CO_2-absorbing regions because these regions comprise IASI channels with weighting functions that peak at similar altitudes. More specifically, the peaks of the weighting functions of the shortwave channels more strongly affected by NLTE emission (i.e., those between 2,240 and 2,385 cm^{-1}) cover a range (~6 to ~400 hPa) similar to that covered by the peaks of the weighting functions of the channels between 645 and 710 cm^{-1}. It should be noted, however, that longwave channels near the center of the Q-branch of the fundamental CO_2 ν_2 band at 667.38 cm^{-1} (i.e., channels between 667.5 and 668.5 cm^{-1}) peak at altitudes (~6 to ~1 hPa) not reached by channels in the 4.3-μm region. We should also stress that, as discussed in section 2, temperature sounding channels in the 15 μm region are negligibly affected by NLTE emission.

The means of the daytime difference statistics, $O - B$, for the longwave and the shortwave IASI channels are shown in the upper and lower panels of Figure 8, respectively. In the left panels we have plotted results for the boreal winter period discussed previously (i.e., 10–20 November 2014), whereas in the right panels we have plotted results for a boreal summer period of equal length (i.e., 10–20 June 2015). The curves plotted in Figure 8 indicate results for IASI soundings between 60°N and 80°N. We have chosen these latitudes because during winter they correspond to solar zenith angles larger than 75°, that is, the angles at which radiance biases become increasingly larger. Note that during summer the same latitudes correspond to solar zenith angles larger than ~40°. The most striking feature of Figure 8 is the difference between the high-latitude shortwave biases in winter and summer. More specifically, Figure 8 shows a dramatic reduction of the biases during summer. If the enhanced winter biases were caused by enhanced biases in the stratospheric/lower-mesospheric temperature fields used as input to the simulations, then we would expect to see a similar pattern in the longwave biases. However, the curves plotted in the top panels of Figure 8 indicate that there is little difference between longwave biases in winter and summer.

With regard to the longwave biases, we note that for the channels comprised in the spectral region discussed in the paper, emission is coming primarily from the stratosphere where the temperature increases with altitude. Hence, the more weakly absorbing channels (between absorption lines) see lower in the atmosphere, whereas the more strongly absorbing channels (on absorption line centers) see higher in the atmosphere. In that respect, it is interesting to note that longwave biases for the more weakly absorbing channels are virtually unchanged between the two seasons, whereas, during summer, biases for the more strongly absorbing channels shift toward more negative values (note that the larger shift is observed for the very-high peaking channels in the Q-branch). Although this behavior might reflect the seasonal evolution of systematic errors in the temperature fields in the upper atmosphere, the magnitude of the effect is still too small to justify the large differences observed in the shortwave. It should also be noted that during summer shortwave biases decrease, whereas longwave biases do actually increase. Thus, the large shortwave biases observed at winter high latitudes are most likely due to the simulation of the NLTE radiances.

As discussed in section 3, the NLTE LBL spectra utilized to train the new fast NLTE model (and, we should add, all fast NLTE models published to date) cover a narrow range of ozone mixing ratios. This is a potential source of error because at altitudes of 50–70 km, O_3 constrains the concentration of the excited oxygen atom $O(^1D)$, which is an important NLTE excitation source of the CO_2 (ν_3 = 1) level. The high-latitude winter conditions discussed in the previous paragraphs correspond to the occurrence of enhanced ozone mixing ratios in the mesosphere. As discussed by, for example, Sofieva et al. (2009), tertiary ozone maxima are observed at

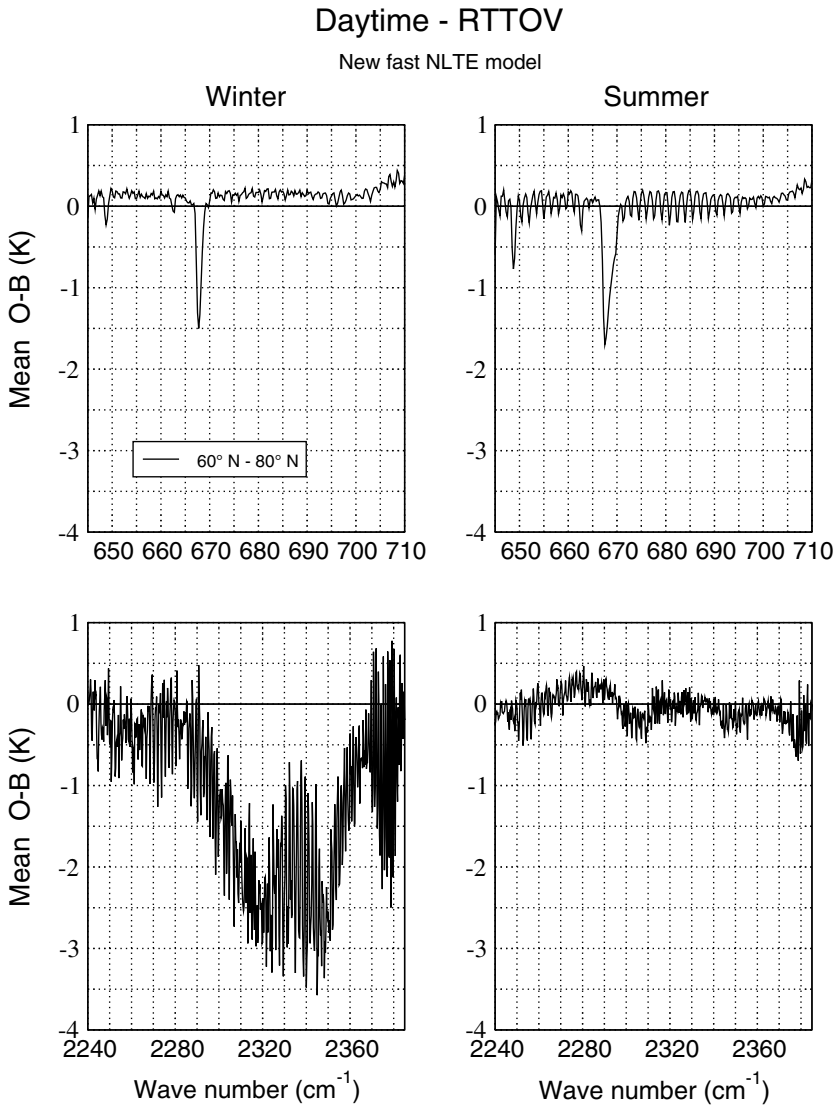

Figure 8. The mean value of the difference between Infrared Atmospheric Sounding Interferometer daytime observations and RTTOV simulations for longwave radiances (upper panels) and shortwave radiances (lower panels). Results are presented for the geographical region comprised between 60°N and 80°N. RTTOV = radiative transfer for TIROS operational vertical sounder; NLTE = nonlocal thermodynamic equilibrium.

high latitudes at the beginning (October–November), in the middle, and at the end (February–March) of winter. Tertiary ozone maxima are caused by the decrease in atomic oxygen losses by catalytic cycles involving the odd-hydrogen species OH and HO_2. The peak of an ozone tertiary maximum can reach 4 ppmv, and its position is between 65 and 75 km depending on latitude and time. An example of ozone tertiary maximum is shown in the lower panel of Figure 9 (i.e., the O_3 profile represented by the black curve). This ozone profile is part of the suite of data belonging to the January-75°N atmosphere in the "IG2" data set by Funke et al. (2012; see Table 2 of Funke et al., 2012, for details). As shown in the lower panel of Figure 9, between 50 and 70 km the IG2 O_3 profile lies well outside the envelope of minimum (the dashed red curve) and maximum (the dashed green curve) values covered by the ozone profiles used in the computation of the NLTE LBL training spectra. Therefore, at winter high latitudes we expect that the enhanced $O(^1D)$ excitation of the CO_2 ($\nu 3 = 1$) level will be poorly represented in the fast NLTE model simulations. More specifically, the fast NLTE model should simulate spectra that are colder than the observed ones because enhanced $O(^1D)$ concentrations produce warmer vibrational temperatures, hence warmer radiances. This is illustrated in the upper panel of Figure 9 where we have plotted the difference

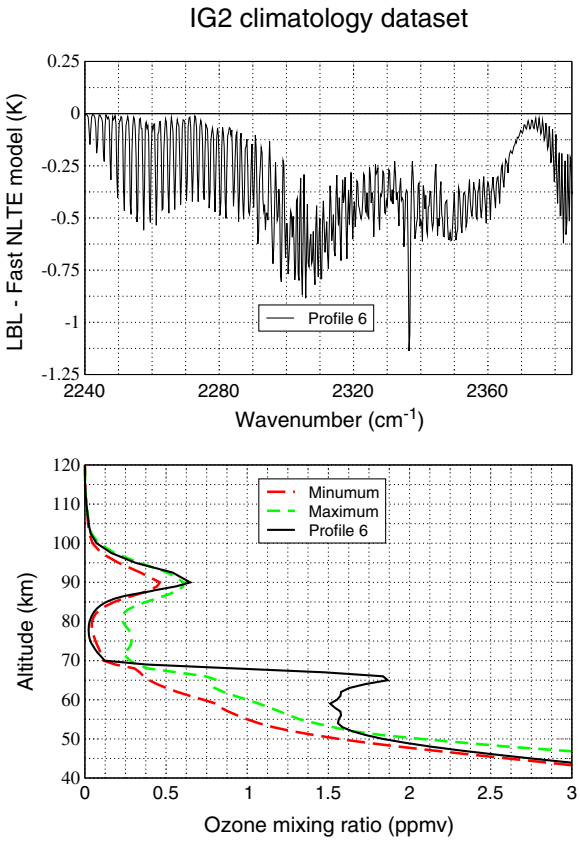

Figure 9. Top panel: The difference between the NLTE line-by-line spectrum corresponding to the IG2 January-75°N atmosphere and the NLTE line-by-line spectrum corresponding to the same atmosphere but with the ozone profile replaced by one of the profiles used for the training of the fast NLTE model. Lower panel: The ozone profile (black curve) for the January-75°N atmosphere. The dashed red and dashed green curves represent the envelope of minimum and maximum ozone values utilized to train the new fast NLTE model. NLTE = nonlocal thermodynamic equilibrium.

between the NLTE LBL spectrum corresponding to the IG2 January-75°N atmosphere and the NLTE LBL spectrum corresponding to the same atmosphere but with the ozone profile replaced by one of the profiles used for the training of the fast NLTE model. Figure 9 shows that although the magnitude of the effect due to the enhanced $O(^1D)$ excitation of the CO_2 (v3 = 1) level is far from negligible, this effect cannot explain the enhanced biases at winter high latitudes because the mean *O-B* departures should have an opposite sign. This said, we should stress that there is still a need to improve the training of the fast NLTE model by introducing an adequate representation of the ozone variability in the mesosphere.

In our view, results presented in this section give convincing evidence that neither biases in stratospheric/lower-mesospheric temperature fields nor ozone effects are responsible for the enhanced daytime shortwave radiance biases at winter high latitudes. With regard to the origin of these biases, a possible explanation is that the pressure/temperature training profiles do not represent adequately the conditions found at high latitudes during winter. There could also be an effect related either to the representativeness of the vibrational temperatures at large solar zenith angles or to the inhomogeneity of the solar illumination along the line of sight (in the lower mesosphere, the vibrational temperatures of CO_2 change rapidly at large solar zenith angles). However, a thorough investigation of these hypotheses would require substantial modifications to the fast NLTE model and this is outside the scope of the paper.

5. The Training of PC-RTTOV in Presence of NLTE Effects

Having assessed the performance of the new fast NLTE model, we can now describe the methods applied to train the PC-RTTOV model in presence of NLTE effects. For a detailed description of PC-RTTOV, the reader can refer to Matricardi (2010).

As briefly discussed in section 3, PC-RTTOV exploits PCA to replace a large number of highly correlated spectral radiances with a smaller number of orthogonal PCs. The PCs are calculated from a dataset that comprises *l* synthetic LBL spectra described by radiances in *n* channels. The training spectra represent the range of variations found in the atmosphere. The radiances to be analyzed by PCA are held in the $n \times l$ matrix **X.** If **C** is the $n \times n$ covariance matrix of the spectra, PCA involves computing the *n* eigenvectors and the *n* eigenvalues of the covariance matrix **C**. The eigenvectors are known as PCs. The PCs represent the directions of maximum variance in the data. If λ_i is the eigenvalue associated with the *i*th eigenvector, then the value of $\lambda_i^2/\Sigma\lambda_i^2$ represents the proportion of variation explained by the *i*th PC. The PCs are orthogonal, hence uncorrelated, and the values of the PCs for a spectrum are known as PC "scores." The PC scores **p** can be interpreted geometrically as the projection of the radiance spectra **y** into the principal components:

$$\mathbf{p}=\mathbf{A}^T\mathbf{y} \tag{3}$$

where **A** is the matrix of PCs (here superscript *T* denotes the transpose). Although there are as many PCs as there are channels, most of the atmospheric variability in the data set can be represented by a number of PCs, *m*, fewer than *n*. We can then reduce the dimension of the dataset by replacing *n* radiances with the first *m* PCs. In many applications, the choice of the number of dimensions is based on the total variation accounted for by the leading PCs and it will in general depend on specific aspects of the data set. Starting from equation (3), the reconstructed full radiance spectrum $\tilde{\mathbf{y}}$ is obtained as follow:

$$\tilde{\mathbf{y}}=\mathbf{A}_m\mathbf{p} \tag{4}$$

where m is the number of truncated eigenvectors. Based on equation (4), one can envisage the development of a PC-based fast radiative transfer model by devising a computationally efficient algorithm that allows the fast simulation of PC scores for spectra included or not included in the PCA.

PC-RTTOV performs rapid and accurate simulations of PC scores of high-spectral-resolution infrared radiances using a multiple linear regression scheme where the PC scores are expressed as a linear combination of predictors that depend on the atmospheric state vector. The predictors employed by the PC-RTTOV fast model are polychromatic TOA radiances computed by the classical RTTOV model. The regression coefficients are obtained by fitting the regression equation to a database of PC scores of synthetic noise-free clear-sky radiances calculated using an accurate LBL model. The PC scores can be computed for any state vector that includes variable profiles of temperature, water vapor, ozone, and surface parameters.

The current version of PC-RTTOV is trained using 15,000 LBL spectra representative of atmospheric situations over land and sea for a range of sensor zenith angles that cover the viewing geometry of IASI (Matricardi, 2015). These spectra have been computed using the LBL model described in section 3. To develop a NLTE version of PC-RTTOV, the basic data set of 15,000 LTE spectra has been augmented with additional spectra obtained by adding a NLTE correction to the LTE radiances. The NLTE correction has been computed using the new fast NLTE model described in section 3. We have applied NLTE corrections corresponding to the 12 solar zenith angles used for the training of the fast NLTE model based on the new GRANADA vibrational temperatures. Thus, the total number of spectra used for the training of the NLTE version of PC-RTTOV amounts to 195,000. For testing purposes, we have generated an additional data set comprising 12,000 LTE spectra and 144,000 NLTE spectra corresponding to atmospheric situations independent of those used to train PC-RTTOV. Note that the independent NLTE spectra have been computed using the same fast NLTE model utilized for the computation of the NLTE training spectra.

5.1. The Radiance Reconstruction Error

As shown by equation (4), any spectrum can be reconstructed to an accuracy that will depend on the number of eigenvectors used in the reconstruction. Thus, the final accuracy of the PC-RTTOV radiance simulations will depend on the number of truncated eigenvectors, LBL/spectroscopy errors, and PC score parameterization errors. If $m \leq n$ is the number of truncated eigenvectors (n is the dimension of the spectrum, i.e., the total number of channels in the spectrum), the reconstruction error for each channel, ε_i, can be computed as follows:

$$\varepsilon_i = \sqrt{\frac{\sum_{j=1}^{l}\left[\sum_{k=1}^{n} A_{k,\ i} p_{k,\ j} - \sum_{k=1}^{m} A_{k,\ i} p_{k,\ j}\right]^2}{l}} \tag{5}$$

In equation (5), l is the number of LBL spectra in the training data set, whereas $A_{k,\ i}$ and $p_{k,\ j}$ represent the eigenvectors and the PC scores of the LBL training spectra, respectively. Note that, as in other sections of the paper, radiances are expressed as brightness temperatures in kelvin units (K).

We have computed the radiance reconstruction error, ε_i, for different values of m for each of the 8,461 channels of IASI. Results are shown in the four panels of Figure 10 where we have plotted reconstruction errors using 150, 200, 300, and 400 eigenvectors, respectively. In each panel, the black curve represents the reconstruction error, whereas the red curve represents the instrument noise. Note that for each channel the instrument noise has been computed at a scene temperature corresponding to a global mean value. For all the cases that we have tested, the reconstruction error is well below the instrument noise for all channels and decreases steeply with the number of truncated eigenvectors.

5.2. The PC Score Statistical Regression Algorithm

The PC-RTTOV model is based on a computationally efficient algorithm that allows the fast simulation of PC scores. This task is performed by a statistical regression algorithm formulated fitting an equation to the LBL PC score training data. As discussed by Matricardi (2010), we fit an equation in the form of a polynomial that is linear in the polychromatic radiances of the channels of the selected sensor. The efficiency of the algorithm hinges on the rapid computation of the polychromatic radiances. In PC-RTTOV, we perform this task using the

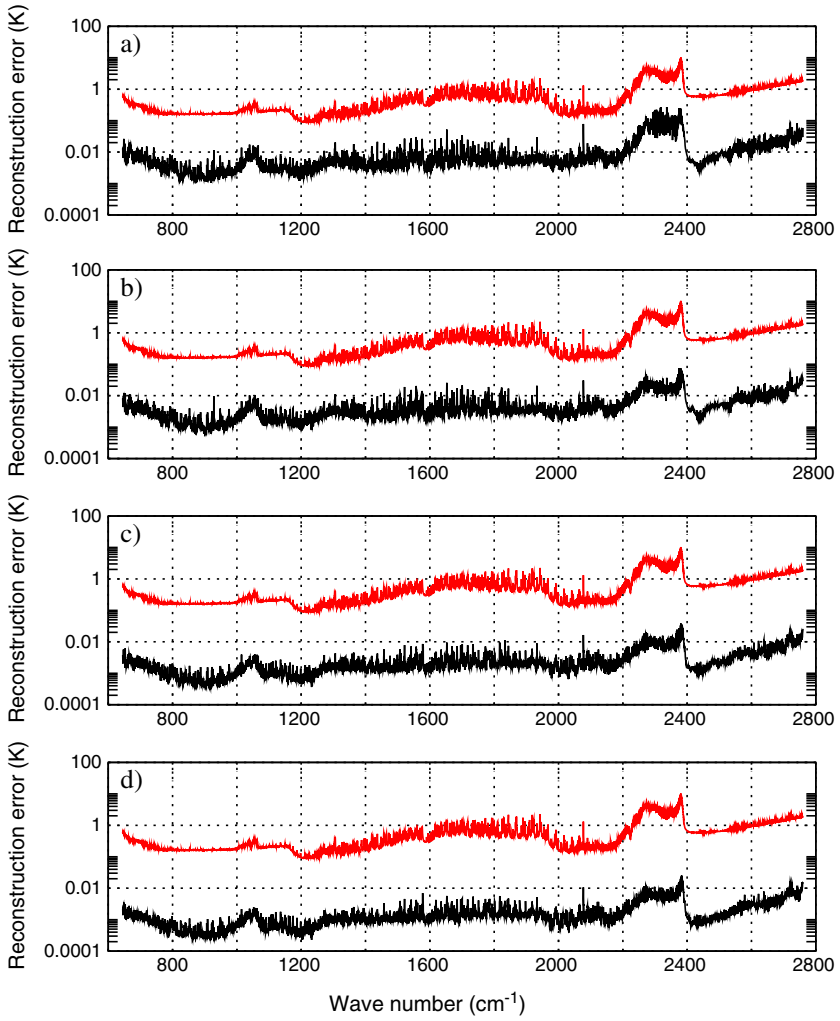

Figure 10. The reconstruction error for the Infrared Atmospheric Sounding Interferometer channels using (a) 150 principal components, (b) 200 principal components, (c) 300 principal components, and (d) 400 principal components. The red line is the instrument noise computed at a scene temperature corresponding to a mean value of the radiance.

classical RTTOV fast radiative transfer model. Our problem then reduces to the computation of coefficient estimates in the multiple linear regression model defined by

$$\mathbf{p}=\mathbf{Rb} \tag{6}$$

where **p** is the vector of l observations of the dependent variable (i.e., the principal component scores of the LBL spectra), **R** is the l by $c \leq n$ matrix of the independent variable (i.e., the RTTOV polychromatic radiances), and **b** is a vector of length c of unknown coefficients that must be estimated.

Because we are trying to establish a statistical relationship between the LBL PC scores and the RTTOV radiances, the absolute accuracy of the RTTOV radiances is not of paramount importance in this process as long as they can reproduce the variability of the LBL radiances in the training data set. Given the regression coefficients **b**, PC-RTTOV can then perform a fast calculation for the PC scores on a completely new atmospheric state vector at any viewing angle within the given range, simply by calculating c new predictor values.

In a linear regression model, the coefficient estimates describe the characteristics of the population from which the sample is taken and are interpreted as a measure of the true characteristics of a population. However, independent variables can be linearly related to another independent variable or to a subset of independent variables and this can result in a value of the estimated coefficients in a sample that can differ significantly from the true value in the population. Consequently, the findings for a sample can be difficult to replicate in an independent sample from the same population (i.e., the coefficient estimate is ill conditioned).

Table 2
The Value of the Correlation Threshold Used in the Selection of the PC-RTTOV Predictors

Number of predictors	Value of the threshold for IASI
300	0.9984
400	0.9987
500	0.9987
600	0.9985

It is then important that for the estimation of the coefficients, the effect of colinear relations between the independent variables is reduced as much as possible.

In our regression model, the independent variables are chosen among the thousands of IASI channel radiances. Many of the IASI channels have a similar atmospheric sensitivity, and, as discussed above, radiances in channels that are similar in content can cause problems in the regression since they can be very highly correlated. To reduce the effect of colinear relations between the independent variables, we have arranged the channels in clusters that contain highly correlated channels. To calculate the clusters, we follow the approach by McMillin and Goldberg (1997). The first step in the calculation of the clusters is represented by the computation of the correlation matrix of the RTTOV training spectra. We then search for the channel with the largest standard deviation (lead channel) and find all the channels that are correlated to the lead channel above a given threshold. The channel cluster thus obtained is removed from the selection pool and the process is repeated until no more clusters can be found. At the end of this process, the first, c, lead channels become the, $c \leq n$, predictors in the regression.

The spectral location of the lead channels and the number and distribution of channels in each cluster depends on the threshold value used to specify the level of correlation. To specify the threshold, we select the value that minimizes the fitting root-mean-square (rms) error defined as follows:

$$\mathrm{rms} = \left[\frac{\sum_{i=1}^{l}\sum_{j=1}^{n}\left(y_{i,j}^{\mathrm{LBL}} - \tilde{y}_{i,j}^{\mathrm{PC}}\right)^2}{l\,n}\right]^{1/2} \tag{7}$$

where $y_{i,j}^{\mathrm{LBL}}$ are the LBL radiances and $\tilde{y}_{i,j}^{\mathrm{PC}}$ are the PC-RTTOV radiances reconstructed using the simulated PC scores. We have evaluated the quantity in equation (7) for various numbers of predictors and threshold values. In particular, we have selected 300, 400, 500, and 600 predictors and have varied the value of the threshold between 0.9970 and 0.9999 in increments of 0.001. Typically, the selected value of the threshold (i.e., the absolute minimum of the fitting error curve) varies slightly with the number of predictors in the regression and does not depend on the truncated number of eigenvectors used to reconstruct the radiances. The threshold values used in this paper are tabulated in Table 2.

Finally, we should stress that the same set of predictors and regression coefficients can be used to simulate either LTE or NLTE radiances. More specifically, if the predictors (i.e., the polychromatic channel radiances) include a NLTE correction, then PC-RTTOV will simulate NLTE radiances in the relevant shortwave spectral region (i.e., 2,200–2,400 cm^{-1}) and LTE radiances in the remaining spectral regions of the IASI spectrum. On the other hand, if the predictors do not include a NLTE correction, then PC-RTTOV will simulate LTE radiances throughout the IASI spectrum. For consistency reasons, the NLTE correction applied to the predictor radiances must be computed by the same NLTE model used in the regression (i.e., the new fast NLTE model).

5.3. Performance of PC-RTTOV for Simulations of IASI Radiances

5.3.1. Validation Against Reference LBL Spectra

In section 5.1 we have discussed the radiance reconstruction error by varying the number of eigenvectors used to reconstruct the radiances. The results shown in Figure 10 were obtained using spectra derived from exact LBL PC scores. We can now use the PC score regression algorithm to assess the accuracy of PC-RTTOV by comparing the radiances computed by the fast model with reference LBL radiances. PC-RTTOV radiances computed for the dependent atmospheric situations used to train the fast model can be compared with the LBL model equivalents to determine the accuracy of the fast model itself. In the same context, we shall use the set of independent atmospheric situations to allow uncertainties from different types of atmospheres to be included.

The PC-RTTOV radiances can be computed using a variable number of predictors and truncated eigenvectors. Since the number of predictors and the number of eigenvectors can be combined in many different ways, it is

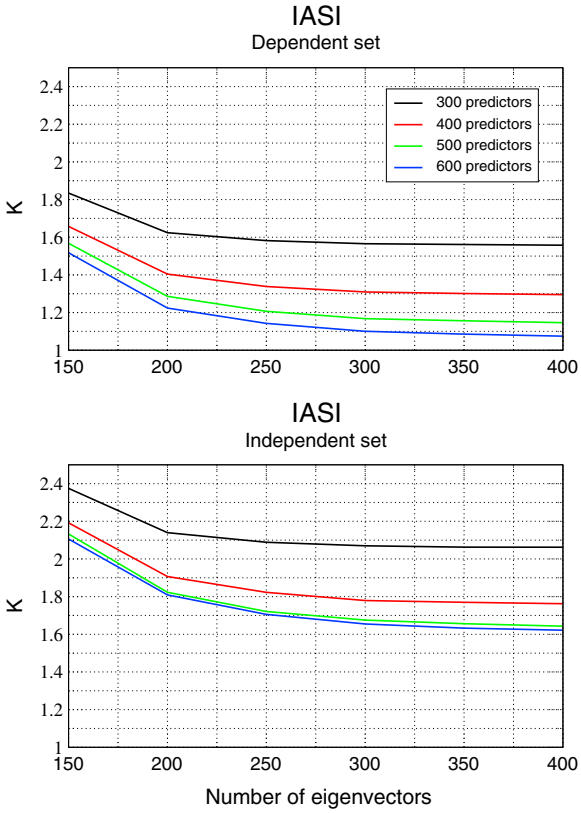

Figure 11. The root-mean-square fitting error as a function of the number of predictors and the number of eigenvectors used in the radiance reconstruction. Results in the upper panel are for the 195,000 spectra used for the training, whereas results in the lower panel are for the 156,000 spectra in the independent data set. IASI = Infrared Atmospheric Sounding Interferometer.

impractical to present results on a single channel basis for all the possible combinations. Consequently, we will first concentrate on the discussion of the results in terms of the rms fitting error computed using equation (7) and then present results on a single channel basis only for a selected number of cases. The rms fitting error is plotted in the upper and lower panel of Figure 11 for the dependent and independent case, respectively. Results are expressed as a function of the number of eigenvectors and the number of predictors. Results for the independent case give a more realistic estimate of the error in that they tell us how good we are in replicating the findings for the training sample in an independent sample from the same population.

The two panels in Figure 11 show that an increase in the number of predictors results in a reduction of the rms fitting error. The dependence of the error on the number of predictors is less accentuated for the independent case suggesting that errors related to the representativeness of the training set are more important than errors determined by the choice of the number of predictors. This is also reflected in the fact that when moving to the independent data set, errors obtained using more predictor variables tend to increase more than errors obtained using less predictor variables, although, it should be said, the effect is very small. It is also evident that when we use fewer eigenvectors for the radiance reconstruction, the total error budget tends to be dominated by the number of eigenvectors used in the radiance reconstruction. As shown in Figure 11, the use of 300 or, indeed, 400 eigenvectors gives little advantage over the use of 250 eigenvectors.

Results for the independent set on a single-channel basis (i.e., in equation (7) we do not perform the summation over the number of channels, *i*) are presented in Figure 12 where we have plotted results for two cases selected by combining the minimum and maximum number of predictors with the maximum allowed number of eigenvectors (i.e., 400). It can be seen that increasing the number of predictors has a tangible effect on the accuracy of the simulations. Although not shown here, this effect is less accentuated when we reduce the number of eigenvectors used for the radiance reconstruction. The choice of number of eigenvectors also dominates the error budget in the shortwave region of the spectrum. It should be noted, however, that, irrespective of the choice of tuning parameters, the rms error is always well below the instrument noise. The bias makes a marginal contribution to the rms error, and for the vast majority of the channels, maximum errors typically do not exceed ±2 K with the noticeable exception of the channels in the shortwave region of the spectrum beyond 2,600 cm^{-1}. We can conclude this section by noting that results shown in Figure 12 are almost identical to those presented by Matricardi (2015). This means that the regression coefficients described in this paper can replace the previous regression coefficients with no loss of accuracy.

5.3.2. Validation Against Observations

In this section, we present results obtained by comparing IASI observations to simulations carried out using the NLTE version of PC-RTTOV. The experimental setup is identical to the one described in section 4.2. For comparison purposes, PC-RTTOV results are presented alongside those obtained using the classical RTTOV model. Both RTTOV and PC-RTTOV incorporate the new fast NLTE model based on the new GRANADA vibrational temperatures. Results for PC-RTTOV are shown in the left panels, whereas results for RTTOV are shown in the right panels. Here we should stress that we are not aiming to compare the absolute performance of the two fast RT models. Although we know that PC-RTTOV can reproduce the underlying LBL radiances much more accurately than the classical RTTOV (see Matricardi, 2010, for details), when comparing observations to simulations, the error budget is dominated by LBL spectroscopic errors and uncertainties in the specification of the input state vector. Rather, results shown in Figure 13 serve as a sanity check to demonstrate that the NLTE version of PC-RTTOV does not give unexpected results. In that respect, Figure 13 clearly shows that PC-RTTOV biases and standard deviations are very similar to those obtained using the classical (and well

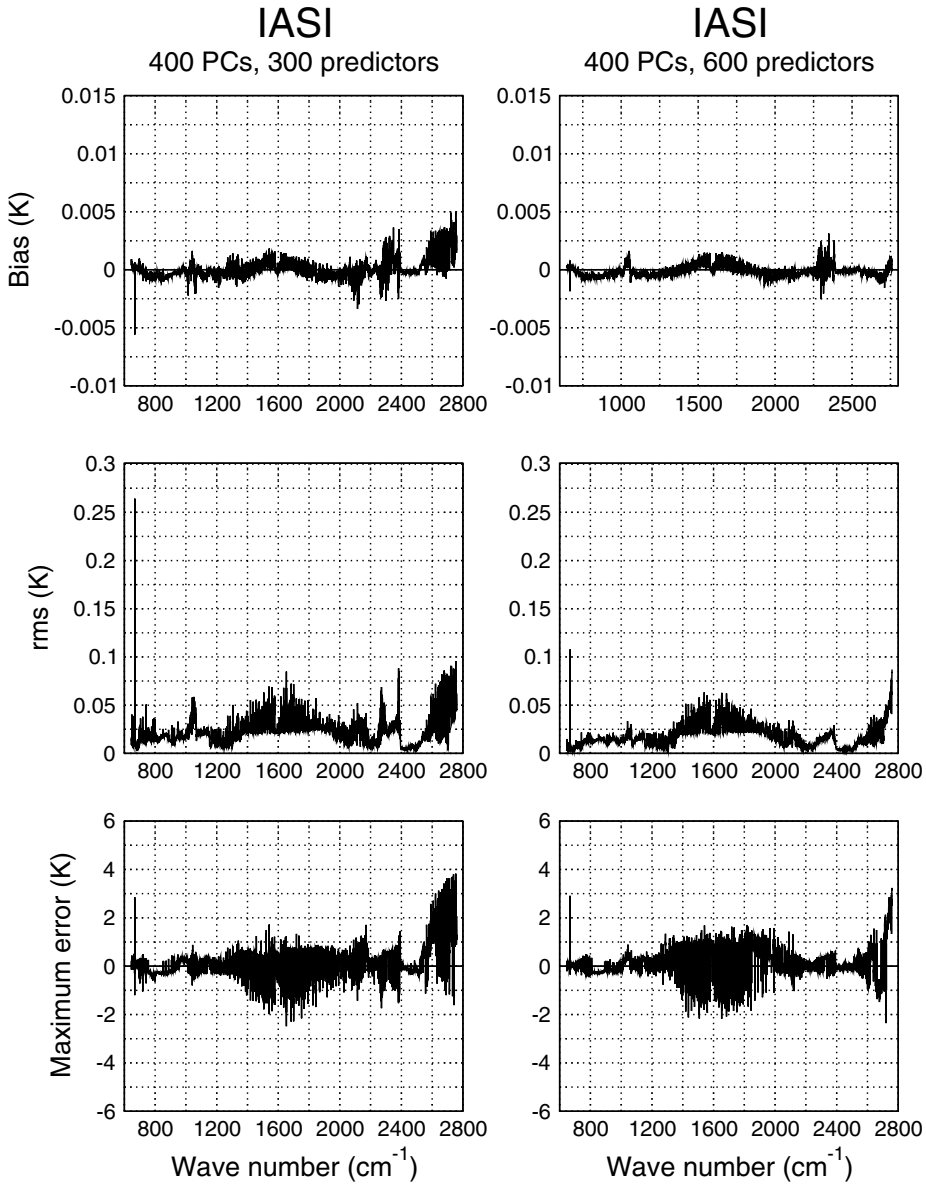

Figure 12. The mean value (bias; upper panel), root-mean-square (middle panel), and maximum value of the difference (lower panel) between PC-radiative transfer for TIROS operational vertical sounder and line-by-line radiances using 400 PCs for the radiance reconstruction. Results have been obtained for the 156,000 spectra in the independent data set. The statistics has been computed using local thermodynamic equilibrium and nonlocal thermodynamic equilibrium spectra. IASI = Infrared Atmospheric Sounding Interferometer; PC = principal component.

established) version of RTTOV. The robustness of the PC score prediction algorithm is further demonstrated in Figure 14, where we have plotted the statistics of daytime *O-B* departures obtained by removing NLTE effects in the simulations.

6. Summary and Future Directions

In this paper, we have documented the introduction of NLTE effects in the PC-RTTOV (Matricardi, 2010) fast radiative transfer model. The primary aim of this development is toward allowing the use of PC-RTTOV for the simulation of IASI daytime shortwave PC scores or reconstructed radiances. The inclusion of NLTE effects in PC-RTTOV has required the development of a parameterized scheme that allows the fast computation of a NLTE correction to LTE radiances. The fast NLTE model is based on the use of state-of-the-art vibrational

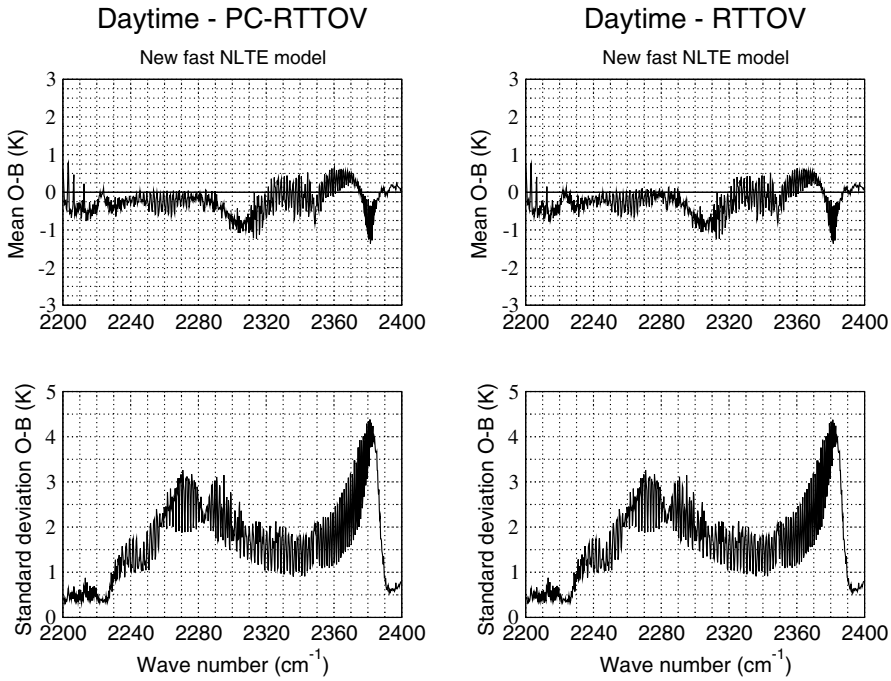

Figure 13. The mean value (bias; upper panels) and the standard deviation (lower panels) of the difference between observed daytime radiances and radiances computed using the PC-RTTOV (left panels) and RTTOV (right panels) model. Simulations include NLTE effects. PC = principal component; RTTOV = radiative transfer for TIROS operational vertical sounder; NLTE = nonlocal thermodynamic equilibrium.

temperatures computed using the latest version of the GRANADA NLTE population algorithm (Funke et al., 2012) and is general enough to be applied to any sensor and can be utilized to add a fast and accurate NLTE correction to the LTE spectra computed by any general polychromatic RT model, including RTTOV,

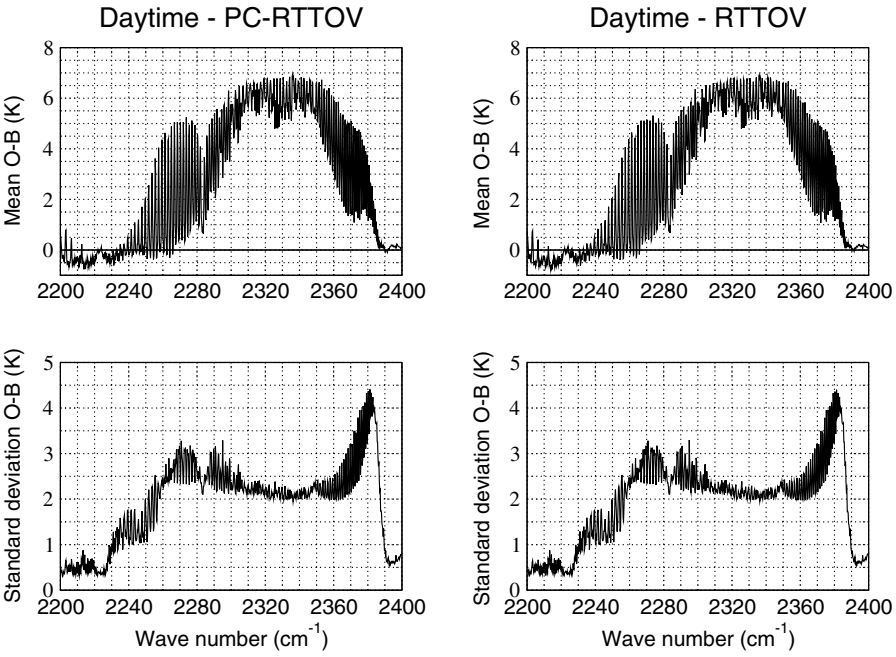

Figure 14. The mean value (bias; upper panels) and the standard deviation (lower panels) of the difference between observed daytime radiances and radiances computed using the PC-RTTOV (left panels) and RTTOV (right panels) model. Simulations do not include nonlocal thermodynamic equilibrium effects. PC = principal component; RTTOV = radiative transfer for TIROS operational vertical sounder.

which now incorporates the fast NLTE model developed in this study. We have extensively tested the fast NLTE model against IASI observations. Testing over a 10-day winter period shows that the computation of the daytime NLTE radiances is almost as accurate as the computation of the nighttime radiances. In addition, we have shown that the fast NLTE model presented in this paper performs significantly better than the fast NLTE model incorporated in the previous version of the RTTOV model (RTTOV-11, Saunders et al., 2013). The fast NLTE model has also been used to carry out trials to study the relative merits of vibrational temperatures computed using different versions of the GRANADA NLTE population algorithm, that is, the old GRANADA algorithm by Funke et al. (2002) and the new GRANADA algorithm by Funke et al. (2012). Results show that in many spectral regions the fast NLTE model based on the old GRANADA vibrational temperatures has a lower bias. However, it is difficult to conclude that the old vibrational temperatures are more accurate than the new ones. Because biases are generally small, there is no obvious way to distinguish between systematic errors from the atmospheric state, the RTTOV forward model (including spectroscopic errors) and the fast NLTE model. The new vibrational temperatures might still be more accurate and yet result in larger residuals due to the competing effect of biases from other sources. These findings warrant further investigation in future studies.

While the performance of the fast NLTE model is overall very good, it is important to highlight its current limitations. Arguably, the most important of these is the accuracy of the NLTE simulations at winter high latitudes where departures from observations are significantly larger than departures at lower latitudes. Although some authors (DeSouza-Machado et al., 2007) have suggested that these larger departures are due to enhanced biases in the stratospheric/lower-mesospheric temperature fields used in the simulations, in this paper we provide convincing evidence that temperature biases are not responsible for the enhanced radiance biases. In the paper we have also investigated the possibility that these biases might be related to the fast NLTE model poor representation of effects due to the presence of ozone tertiary maxima in the mesosphere at high latitudes during winter. Results show that although there is a need to improve the training of the fast NLTE model by introducing an adequate representation of the ozone variability in the mesosphere, enhanced ozone concentrations in the mesosphere cannot explain the larger biases at high latitudes. With regard to the origin of these biases, a possible explanation is that the pressure/temperature training profiles do not represent adequately the conditions found at high latitudes during winter. There could also be an effect related either to the representativeness of the vibrational temperatures at large solar zenith angles or to the inhomogeneity of the solar illumination along the line of sight. However, a thorough investigation of these hypotheses would require substantial modifications to the fast NLTE model and this is outside the scope of the paper. We plan to address these issues in a future phase of the study.

Finally, the comparison of IASI observations to simulations carried out using the NLTE version of PC-RTTOV shows that biases and standard deviations are very similar to those obtained using the classical (and well established) version of RTTOV. This result holds irrespective of whether simulations include or not include NLTE effects, thus attesting to the robustness of the PC score prediction algorithm.

Acknowledgments

We thank Steve English (ECMWF) for his comments on the original version of the paper and Sergio DeSouza-Machado (UMBC) and two anonymous reviewers for their effort in helping improving the quality of the paper. Marco Matricardi was jointly supported by ECMWF and EUMETSAT. M. Lopez-Puertas and B. Funke were supported by the Spanish MICINN under grant ESP2014-54362-P and EC FEDER funds. The Line-by-line Radiative Transfer Model (LBLRTM) used to compute the training data set of LTE and NLTE radiances is publicly available from the atmospheric and environmental research (AER) Radiative Transfer Working Group Website (http://rtweb.aer.com/lblrtm_frame.html). The vibrational temperature data sets can be accessed at http://www.iaa.es/~puertas/jgr_paper_2017.html. Finally, the ECMWF geophysical profiles used to compute the statistics of the *O-B* departures can be retrieved from the ECMWF archive by contacting ECMWF data services at data.services@ecmwf.int. For more information on ECMWF data services, the reader can consult the following web page: https://www.ecmwf.int/en/forecasts/accessing-forecasts.

References

Antonelli, P., Revercomb, H. E., Sromovsky, L. A., Smith, W. L., Knuteson, R. O., Tobin, D. C., et al. (2004). A principal component noise filter for high spectral resolution infrared measurements. *Journal of Geophysical Research*, *109*, D23102. https://doi.org/10.1029/2004JD004862

Chalon, G., Cayla, F., & Diebel, D. (2001). IASI: An advanced sounder for operational meteorology. In *Proceedings of the 52th Congress of IAF*, (pp. 1–5). Tolouse, France.

Chen, Y., Han, Y., van Delst, P., & Weng, F. (2013). Assessment of shortwave infrared sea surface reflection and nonlocal thermodynamic equilibrium effects in the community radiative transfer model using IASI data. *Journal of Atmospheric and Oceanic Technology*, *30*(9), 2152–2160. https://doi.org/10.1175/JTECH-D-12-00267.1

Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., et al. (2005). Atmospheric radiative transfer modeling: A summary of the AER codes, short communication. *Journal of Quantitative Spectroscopy & Radiative Transfer*, *91*(2), 233–244. https://doi.org/10.1016/j.jqsrt.2004.05.058

Collard, A. D., & McNally, A. P. (2009). Assimilation of Infrared Atmospheric Sounding Interferometer radiances at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, *135*, 1036–1050.

Collard, A. D., McNally, A. P., Hilton, F. I., Healy, S. B., & Atkinson, N. C. (2010). The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *136*(653), 2038–2050. https://doi.org/10.1002/qj.701

DeSouza-Machado, S., Strow, L. L., Hannon, S. E., Motteler, H. E., López-Puertas, M., Funke, B., & Edwards, D. P. (2007). Fast forward radiative transfer modeling of 4.3 μm nonlocal thermodynamic equilibrium effects for infrared temperature sounders. *Geophysical Research Letters*, *34*, 01802. https://doi.org/10.1029/2006GL026684

Edwards, D. P., López-Puertas, M., & López-Valverde, M. A. (1993). Non-local thermodynamic equilibrium studies of the 15-μm bands of CO_2 for atmospheric remote sensing. *Journal of Geophysical Research*, *98*(D8), 14,955–14,977. https://doi.org/10.1029/93JD01297

Funke, B., López-Puertas, M., García-Comas, M., Kaufmann, M., Höpfner, M., & Stiller, G. P. (2012). GRANADA: A Generic RAdiative traNsfer AnD non-LTE population algorithm. *Journal of Quantitative Spectroscopy & Radiative Transfer*, *113*(14), 1771–1817. https://doi.org/10.1016/j.jqsrt.2012.05.001

Funke, B., Martín-Torres, F. J., López-Puertas, M., Höpfner, M., Hase, F., López-Valverde, M. A., & García-Comas, M. (2002). A generic non–LTE population model for MIPAS–ENVISAT data analysis, vol. 4, Abstracts of the Contributions of the European Geophysical Society, Nice, France, 21–26 April 2002, CD-ROM, ISSN:1029–7006.

Havemann, S. (2006). The development of a fast radiative transfer model based on an empirical orthogonal functions (EOD) technique. Proc. SPIE 6405, Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications, Goa, India. https://doi.org/10.1117/12.693995

Huang, H.-L., Smith, W. L., & Woolf, H. M. (1992). Vertical resolution and accuracy of atmospheric infrared sounding spectrometers. *Journal of Applied Meteorology*, *31*, 265–274.

Joliffe, I. T. (2002). *Principal component analysis*. New York: Springer.

Jurado-Navarro, Á. A., López-Puertas, M., Funke, B., García-Comas, B., Gardini, A., Stiller, G. P., & Clarmann, T. V. (2015). Vibrational-vibrational and vibrational-thermal energy transfers of CO_2 with N_2 from MIPAS high-resolution limb spectra. *Journal of Geophysical Research: Atmospheres*, *120*, 8002–8022. https://doi.org/10.1002/2015JD023429

Jurado-Navarro, A. A., López-Puertas, M., Funke, B., Garcia-Comas, M., Gardini, A., González-Galindo, F., et al. (2016). Global distributions of CO_2 volume mixing ratio in the middle and upper atmosphere from daytime MIPAS high-resolution spectra. *Atmospheric Measurement Techniques*, *9*(12), 6081–6100. https://doi.org/10.5194/amt-9-6081-2016

Liu, X., Smith, W. L., Zhou, D. K., & Larar, A. (2006). Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept. *Applied Optics*, *45*, 201–209.

López-Puertas, M., & Funke, B. (2015). Non-local thermodynamic equilibrium. In G. R. North, J. Pyle, & F. Zhang (Eds.), *Encyclopedia of atmospheric sciences* (2nd ed., Vol. 5, pp. 16–26). Elsevier. https://doi.org/10.1016/B978-0-12-382225-3.00339-X

López-Puertas, M., & Taylor, F. W. (2001). *Non-LTE radiative transfer in the atmosphere*. Singapore: World Scientific Pub. https://doi.org/10.1142/4650

Matricardi, M. (2010). A principal component based version of the RTTOV fast radiative transfer model. *Quarterly Journal of the Royal Meteorological Society*, *136*(652), 1823–1835. https://doi.org/10.1002/qj.680

Matricardi, M. (2015). The direct assimilation of principal componentes of IASI spectra in the ECMWF 4D-Var and the training of PC_RTTOV over land surfaces, EUMETSAT Contract No. EUM/CO/07/4600001011/OS.

Matricardi, M. (2016). The use of Principal Component techniques to assimilate the full IASI spectrum and the introduction of non-LTE effects in PC-RTTOV, EUMETSAT Contract No. EUM/CO/14/4600001332/TA.

Matricardi, M., Chevallier, F., Kelly, G., & Thepaut, J.-N. (2004). An improved general fast radiative transfer model for the assimilation of radiance observations. *Quarterly Journal of the Royal Meteorological Society*, *130*(596), 153–173. https://doi.org/10.1256/qj.02.181

Matricardi, M., & McNally, A. P. (2013). The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 573–582. https://doi.org/10.1002/qj.2156

McMillin, L. M., & Goldberg, M. D. (1997). The use of super channels for high resolution soundings. In *Proceedings of the Ninth International TOVS Conference* (pp. 20–26). Igls, Austria.

McNally, A. P., & Watts, P. D. (2003). A cloud detection algorithm for high-spectral-resolution infrared sounders. *Quarterly Journal of the Royal Meteorological Society*, *129*, 3144–3423.

Rayer, P. J. (1995). Fast transmittance model for satellite sounding. *Applied Optics*, *34*(31), 7387–7394. https://doi.org/10.1364/AO.34.007387

Rothman, L. S., Gordon, I. E., Barbe, A., ChrisBenner, D., Bernath, P. F., Birk, M., et al. (2009). The HITRAN 2008 molecular spectroscopic database. *Journal of Quantitative Spectroscopy & Radiative Transfer*, *110*(9-10), 533–572. https://doi.org/10.1016/j.jqsrt.2009.02.013

Saunders, R., Hocking, J., Rundle, D., Rayer, P., Matricardi, M., Geer, A., et al. (2013). RTTOV-11 science and validation report, NWP SAF Doc. NWPSAF-MO-TV-032.

Sofieva, V. F., Kyrölä, E., Verronen, P. T., Seppälä, A., Tamminen, J., Marsh, D. R., et al. (2009). Spatio-temporal observations of the tertiary ozone maximum. *Atmospheric Chemistry and Physics*, *9*(13), 4439–4445. https://doi.org/10.5194/acp-9-4439-2009